Automated Multimodal Classroom Interaction Analysis based on Flanders' and Galloway's Interaction Models

Synopsis
Submitted in the partial fulfilment for the requirement of
Doctor of Philosophy
In
Computer Science

Submitted by
Sadhana Singh

Under the supervision of

Prof. Nandita Satsangee
Co-supervisor
Coordinator
School of Education
Dean- Faculty of Education

Dr. Lotika Singh
Supervisor
Dept. of Physics & Computer Science
Faculty of Science

DEPARTMENT OF PHYSICS & COMPUTER SCIENCE,
FACULTY OF SCIENCE
DAYALBAGH EDUCATIONAL INSTITUTE (DEEMED UNIVERSITY)
DAYALBAGH, AGRA (UP) –282005

2021

1. Introduction:

Interaction analysis is the interdisciplinary process for empirical investigation of interaction among the humans and their environment [7]. Interaction analysis investigates the activities of participants involved in an interaction, such as talk, verbal interaction, and non-verbal interaction, behavioral and environmental aspects [2, 11]. Interaction analysis in a classroom is a mechanism for understanding and analyzing the effectiveness of teaching and learning process and the flow of classroom behavior.

In a classroom, teaching and learning is an interactive process that incorporates communication between a teacher and a student. As part of the interaction, a teacher is expected to communicate information, ideas, attitudes and behaviors etc to the students and the students are also expected to participate actively in the interaction process of teaching and learning.

Education Commission (1964-66) has emphasized the significant importance of teacher education programs to improve the effective classroom teaching and learning process. National Education Policy (2019) envisions transforming the nation sustainably into an equitable and vibrant knowledge society. The policy also emphasizes Teachers as the Torchbearers of Change, Teacher Education and integrating Education Technology into all levels of education [44]. To understand the classroom environment and provide qualitative and quantitative feedback to the teachers, various methods have been developed. Classroom Interaction Analysis [1,7] is one of the feedback techniques like Micro teaching [43], Simulated teaching [41] and T-group training [42] for improving the effectiveness of teaching learning processes.

The outcome of the interaction analysis may be useful for improving some of the following aspects of the teaching and learning process:

- More effective verbal and non verbal interaction in the classroom.
- Increasing the Students' participation by the teacher [2,11].
- Positive behavioral changes in teacher's approach.

To conduct a qualitative interaction analysis it is important to capture every dimension of the verbal and non verbal interaction exhibited by the teacher and the student in the classroom. The verbal communication such as elements of speech along with non-verbal communication such as tone, speed, volume, body language, body movements, eye contact, facial expressions and visual cues

demonstrated by the teachers and the students defines the classroom interaction attributes for performing classroom interaction analysis.

To conduct a qualitative interaction analysis of every classroom interaction becomes very difficult as every classroom needs an observer to note and evaluate every aspect of the interaction without any observer biases. The emergence of advancements in technologies such as Artificial Intelligence, Machine Learning, Natural Language Processing and Computer Vision have contributed in the automatic analysis of such complex multimodal behavior in different business applications such as healthcare [11], social web analytics [11], education research [9] and learning analytics[18] etc. But based on our research study, such kind of automated multimodal analysis is not fully explored in education systems particularly for classroom interaction analysis.

Many researchers have proposed different automation tools and advanced learning analytical models [11, 18] to understand and evaluate program education learning outcomes [8]. Some of the researchers have also developed some products for attention monitoring of the students in the classroom [18], but the automation for evaluation of teaching and learning process exhibited in the classroom is not fully explored. This kind of evaluation requires the automatic analysis of teacher and student's classroom interaction by automatically capturing verbal and non-verbal behavioral attributes of teacher and the student without any manual intervention. The focus of this research proposal is to design and develop a qualitative framework and an automated multimodal classroom interaction analysis by capturing all interaction attributes from verbal and non-verbal communication in a classroom environment.

2. Literature Review:

The quality of the classroom interaction between the teacher and student defines the effectiveness of the teaching and learning process in the classroom. The teacher plays dominant role in the classroom interaction because he/she uses different types of communication techniques during the teaching and learning process such as giving directions, asking questions, lecturing, repeating student's response verbatim, and praising and encouraging [1]. A teacher plays an important role in defining the effectiveness of teaching and learning process in a classroom. To understand the effectiveness of teaching learning process in the classroom, it is important to understand the teacher behavior in the classroom and support him with qualitative and quantitative feedback to improve his/her performance

in the classroom. Many researchers have developed different mechanisms and tools to understand teacher's behavior in the classroom. Some of the important tools and techniques designed to observe the teacher's behavior in the classroom are as follows:

- Simulated social skill teaching [41]
- Micro teaching [43]
- Programmed instruction [42]
- T. Group training [42]
- Interaction analysis [5]

Cruick Shank (1968) has developed simulation technique, which is used to generate a set of behaviors in artificial situation. A trainee teacher plays many roles as a teacher, as a pupil and as a mentor [41]. Micro teaching is an effective tool for development of teaching skills. It is originated in Stanford University in 1963. It is similar to simulated social skill teaching that is used to give feedback for the modification of teacher behavior of trainee teacher [43]. The programmed instruction [42] is a highly individualized instructional strategy for the observation and modification of teacher behavior. This technique provides the deep insight and understanding about the elements structure and their sequence of instructional events [42]. T-group training [42] is also an observation technique which is used for trainee teacher behavior modification. It has been developed by Bethel and Mine in 1947. In this a leaders group meets without any pre-planned agenda or schedule [42].

The interaction analysis is another technique for analyzing and observing the classroom behavior [5]. There are different interaction analysis techniques that are used for classroom interaction analysis based on verbal and non-verbal communication strategies. Flander's interaction analysis [4], and Brown's interaction analysis [43] are mainly designed for verbal interaction characteristics. Flander's interaction analysis is considered as one of the most popular verbal interaction analyses, which is developed on a ten category scale for observing the classroom verbal interaction. Galloway's interaction analysis [39] augments the Flanders' interaction analysis model by integrating non-verbal communication aspects of the classroom.

2.1 Classroom Interaction Analysis:

Classroom interaction analysis is a technique for systematic observation and analysis of class environment, teacher behavior and interaction going on in the class. It helps a teacher to modify his teaching traits, techniques, interaction patterns for making his/her teaching more effective and efficient.

"Teachers should include both verbal and non-verbal languages. Verbal interaction covers written interaction and oral interaction". Whereas, non-verbal interaction refers to behavioral responses in classroom without using words such as head-nodding, hand-raising, and so on [1].

According to Dr. S.K. Thakur, class-room interaction analysis "may be defined as an instrument which is designed to record categories of verbal interaction during, or from, recorded teaching learning sessions. It is a technique for capturing qualitative and quantitative dimensions of teacher's verbal behavior in the class-room."

According to Dr. R.A. Sharma "interaction analysis is specialized research procedure that provides information about only a few of many aspects of teaching. It is an analysis of spontaneous communication between teacher and pupils. Interaction analysis applies only to the content free characteristics of verbal communication. The entire process of interaction analysis becomes a measure of teacher influence because it makes the assumption that most teacher influence is expressed through verbal statements and most non-verbal influence is positively correlated with verbal."

2.1.1 Verbal interaction:

Verbal interaction can be described as the message exchanged verbally. It contains messages, information, instructions or directions conveyed with the help of words or text. Interaction Analysis is used as a technique capturing qualitative and quantitative dimensions of teacher's verbal behavior in the classroom [5]. There are several methods of verbal classroom interaction analysis. The popular classroom interaction analysis methods are Flanders' Interaction Analysis Categories (FIAC) [4] and Brown Interaction Analysis System (BIAS) [43]. Flanders developed a tool to study the classroom influence of a teacher on students by observing the only verbal behavior. Brown has also developed a different model to study verbal attributes of the teacher behavior. In both of these models, the observer follows a set of guidelines defined by the respective model to limit the observer bias.

The Flanders System of Interaction Analysis [4] is an observational technique that analyzes only one aspect of spontaneous teacher-student behavior and encodes the behavior in to different categories. It is concerned with verbal behavior only, because verbal behavior can be observed with higher reliability than can nonverbal behavior [5]. Teachers can modify or improve their teaching skills by evaluating the observed information, and can also analyze their own consistency with their intentions and goals for effective decision making in their teaching. Flanders interaction analysis can also help to retrieve the objective information of a teacher's behavior which is abundant to make guidelines for self-directed improvements.

Flanders Interaction Analysis Categories (FIAC) is a Ten Category System of communication which are said to be inclusive of all communication possibilities [5]. These 10 categories of class room interaction are grouped in 3 major categories: Teacher talk, Student talk, and Silence or Confusion. Teacher talk has 7 sub-categories; Student talk has 2 sub categories, and silence.

Table 1. Interaction categories based on Flanders analysis [4].

		Category No.	**Activity**	**Description**
Teacher talk	Indirect influence	1	Accepts feelings	Teacher understands pupils' feelings.
		2	Praises or encourages	Teacher praises pupils' actions or behavior.
		3	Accepts or uses ideas of students	Clarifies or builds or develops ideas suggested by the students.
		4	Asks question	Teacher asks questions.
	Direct influence	5	Lecturing	Teacher explains facts about content.
		6	Gives directions	Teacher gives information, opinions, or orientation.
		7	Criticizes or justifies authority	Statements intended to change students' behavior from non-acceptable to acceptable pattern
Student talk	Response	8	Pupil's Response	Teacher asks questions and students respond.
	Initiation	9	Student- Talk- Initiation	Student asks questions
silence		10	Silence or confusion	Pauses, short periods of silence or confusion

Procedure for Flanders interaction analysis is as follows:

- The observer sits behind the class and observes the classroom environment.

- In every 3 secs the observer assigns a category to that 3 sec slot and this process is continued at a 20 to 25 observation for every minute.
- The observer records the interaction of teacher and student in the Flanders matrix.
- This matrix is prepared for recording the classroom interaction. Flanders Matrix is a 10x10 matrix which consists of all 10 categories mapped with all 10 categories.
- The observer records every pair of observations in the corresponding cells of the matrix. The first observation of every pair maps to row and second observation maps to column of the matrix. Total number of records in the matrix is less than one to the total number of observations.

Flanders suggests a set of ground rules [17] to be followed while noting down the observations in classroom. Some of them are given below.

1. When it is uncertain to place a statement on one of any two categories, choose a category farthest from category five, with the exception of category ten.
2. If the teacher's behavior is either consistently direct or indirect, avoid shifting from one category to the other unless teacher is indicated a clear shift.
3. When the teacher repeats student's answer, if it is a correct answer, then record it as category 2.
4. When a number of students respond to a question, then it is recorded as category 8.

Table 2. Flanders matrix representation [17]

category	1	2	3	4	5	6	7	8	9	10	total
1											
2											
3											
4											
5											
6											
7											

8											
9											
10											
total											

The entries in this matrix are used to find interaction measures defined in the Flanders Interaction Analysis. Interpretation of matrix gives the following different measures as the inferences from it. Some of the measures are:

1. The proportion of teacher talk, student talk and silence.
2. The ratio of indirect and direct influence.
3. Ratio of positive and negative reinforcement.
4. Students' participation ratio.
5. Steady state ratio.
6. Content cross ratio.
7. Constructive integration and vicious cells.
8. Pupil initiation ratio
9. Teacher response ratio
10. Teacher question ratio

An example of classroom interaction in which teacher activities are categorized, is shown below:

Table 3. Classroom interaction instance 1 [7]

Participant	**Actual Classroom Verbal Interaction**	**Code**
T	A thousand pieces of paper cranes. **Do you know paper cranes?**	5 **4**
SS	Origami	
T	Yes! That's origami, colorful paper. **Paper cranes as a gift, you know gift?**	3 **4**
SS	Hadiaaahh…	

In the above example, question was asked in order to assist students getting the notion of a text. The teacher asked students by saying "Do you know paper cranes?" in which account the close-ended questions as students answered with a short phrase "origami." This type of questions is often being

administered by the teacher as he always trying to ensure students understand each difficult word in the text [7].

Another important verbal interaction analysis model is Brown Interaction Analysis System (BIAS) [43]. It developed by Brown and designed for use by teams of students and lecturers in microteaching [43]. This system is simpler than FIAC, with only seven categories, three types of lecturer talk, two of student talk, one silence, and unclassified. In BIAS system, a time-line display sheet is prepared that has three seconds of time quantum for observation.

2.1.2 Non-verbal communication:

Non–verbal interaction is the message or behavior exchanged without words. It can be conveyed with symbols, non-symbols, and expressions, eye-contact, body postures, gestures, touch, conversational silences etc. In a classroom when a teacher is giving directions to students, his/her non-verbal behavior will also be effectively used for instructing the students.

Charles Galloway [3] system of interaction analysis allows the observer to use the same categories, time intervals, and ground rules of Flanders system along with the non-verbal dimensional attributes of teaching [39].

As a part of non-verbal dimensional attributes we may include physical proximity, posture, touch, eye-contact, hand movement and face expressions. Nonverbal cues may also be calculatingly managed to achieve a desired effect, to effect an impression, or to convey an attitude [38].

For example, “Standing with arms folded across the chest psychologically removes the teacher from the class activity. It effectively lets the class know that the teacher will return when certain prerequisites- silence, cleared desk tops, student attention-are met”. Teacher’s eye-contact with students is often used to manage the discipline of the classes. Student communication may be effectively inhibited through stares, eye-rolls, thrown hands, shrugs, and the glances that overlook or see around and beyond the student [3]. Face-to-face verbal communications are always accompanied by non-verbal cues. By the fact of physical presence, this situation cannot be avoided [39].

This research study considers to include important non-verbal features of teacher centric activities like lecturing, giving directions, and teaching through board, teacher movement in the class, body-posture of teacher, gestures and facial expressions of teacher. For student centric activities we will consider a

set of non-verbal features such as face expressions, hand-movements, hand-gestures that can be mapped with the verbal features of interactions described in Flanders system.

2.2 Multimodal Analysis in Classroom Interaction:

Multimodal analysis is considered as a process of encoding, integrating and evaluating the contextual data acquired from multiple input sources for developing a comprehensive contextual understanding. If the multimodal analysis applied in the context of learning, it can be termed as multimodal learning analytics. Multimodal learning analytics can be described as set of approaches that can be used to collect high-frequency data from multiple sources (such as video, logs, audio, gestures, biosensors), then integrate and encode the data, and evaluating learning in realistic mixed-media learning domain [11]. Spikol et al.,[10] defines Multimodal Learning Analytics (MMLA) as a means to capture, integrate and analyze learning traces from different sources to achieve a more holistic and comprehensive understanding of the learning process. MMLA are collected through different kinds of technologies and sensors, and then analyzed with different machine learning techniques [10]. Multimodal Interaction in a classroom environment can be considered as part of Multimodal Learning Analytics. Researchers have also evolved from application of simple machine learning algorithms to deep and recurrent neural networks to process data for multimodal analytics [19]. Boz and Kose [16] used a cascaded feed-forward neural network trained with vortex optimization algorithm. Vortex Optimization Algorithm is an intelligent, Swarm Intelligence based algorithm based on the dynamics of vortices in the nature [16]. Nerina Reitano [13] describes that facial or vocal emotional cues that are not consciously attended to, still influence the processing of facial or vocal stimuli that are attended. A person may respond with respect to the facial expression of a person speaking whose words are coherent with the expressions. Affective computing is a research area to use human emotions to communicate and interact to machines in a realistic and efficient manner. An Affective Computing system examines the detection of facial expressions, speech, body movements, and physiological information. Technologies associated with Affective Computing can be highly useful in building automotive qualitative Multimodal interaction analysis systems.

For conducting the multimodal analysis of classroom interaction, it is important to capture the audio-visual features like speech, text, linguistic, verbal and non-verbal features and behaviors in the

classroom. Recently, researchers have developed systems to capture various non verbal features integrating computer vision techniques to facilitate the multimodal analysis. Some of such systems include facial recognition system CERT [29] which can extract Action Unit features for upper face and lower face features in every frames, MODEC algorithm [22] which can extract the body movement features and recognize upper body gestures. Some other tools like OpenSMILE [34] for extracting the Prosodic, spectral, and voice quality features audio processing.

In face-to-face classrooms scenario, as it is the most general case, it is important to understand what the teacher actually did, as the starting point for teacher reflection and inquiry. Presently, such a kind of teacher enactment recognition and characterization requires costly manual coding by the observer sitting behind the class [12].

Realizing the importance of classroom interaction analysis and difficulty in conducting the manual interaction analysis, Donnelly et al., [14] proposed automated teacher modeling that describes teachers' teaching activities and actions from audio data collected in live classrooms and the system is also evaluated for the automatic teacher analysis consisting of activities or action from audio recordings of classrooms. Features for audio analysis are derived from natural language processing of automatic speech recognition transcriptions and non-verbal acoustic features extracted from noisy classroom audio recordings. Donnelly et al., [14] trained supervised machine learning models to identify occurrences of five key instructional segments Question & Answer, Procedures and Directions, Supervised Seatwork, Small Group Work, and Lecture. These analysis models have note modeled the widely accepted parameters defined in Flanders interaction models or Browns interaction models. Automated teacher modeling [14] divided the recording of each class session into consecutive non-overlapping windows for classification purpose. They evaluated non-overlapping windows for every moment of class time only within a single windowed instance as to not bias the results through classification. Each window was assigned a label of the classroom activity using the segment annotations provided by the classroom coders. Donelly et al., [15] also investigated the detection of teacher questions. Donnelly et al., [14] did not considered to compare the standard education domain research interactional models outcome with the automation model to see the effectiveness of the automation model designed.

Some of the other multimodal systems aligned to teaching learning process include Federico et al., [18] which presents a device for students to video-capture their interactions, poses, emotions and notes by means of a Multimodal Recording Device (MRD). As the lecture completed, the annotations and teacher's slides are combined with the students' notes and the student can get updated notes with the teacher notes and the board content. Every student who attended the lecture gets an enhanced set of notes and lectures. This may help in analysis student attention performance, student behavior. It only provides the data to the students to analyze the details by themselves, there is no automatic analysis report for students.

Considering robot as a teacher, M. Obaid et al., [8] explored an empathy-enabled robotic Tutor with interaction quality, perceived or actual learning effects and the system is based on Emote model. This Interaction analysis updates the learner model with an estimate of the student's valence and arousal during the interaction.

Prieto et al., [12] presented a case study to automatically extract teaching actions during classroom enactment; from five different data stream sources collected using wearable sensors for exploring the potential of different machine learning techniques. Prieto et al., [12] used mobile eye-tracking, EEG, audio-visual and accelerometer data. Though the model uses low-level features, because of integrating wearable sensors, its practicability in implementing a scalable system for classroom discussion becomes debatable.

Kim et al., [9] proposed a feasibility of the system that is capable of making real-time suggestions to an in-class presenter to improve the quality and memorability of their presentation by allowing the presenter to make real-time adjustments/corrections to their non-verbal behavior, such as hand gestures, facial expressions and body language. The proposed system considered to integrate some of the state of the art technologies to measure, electro-dermal activation using physiological sensors, arousal and anxiety of the presented using galvanic skin response, attention using eye tracking devices etc. The actual implementation of scalable system needs to address challenges like 1) the integration of these technologies into a holistic system design 2) their algorithmic adaptation to allow real-time execution and 3) quantification of valid educational variables for use in algorithms [9]. Even if the

system is realized, adopting it in a regular classroom environment is highly debatable as equipping the teacher with such wearable devices and integration of complex technologies.

Considering the above observations and excluding all the complex integrations and wearable sensors, in this research proposal we consider to develop a scalable automated multimodal classroom interaction analysis system by capturing interaction attributes from verbal and non-verbal communication in a classroom environment.

3. Methodology

Classroom interaction analysis requires capturing the verbal and non-verbal features of the classroom interaction, which includes features of teacher and students. For verbal interaction analysis, this research study considers to use a subset or all the features described in the Flanders interaction analysis system. For non-verbal interaction analysis this study considers a set of features aligned and correlated to verbal interaction. For example, for category 1 i.e. accept feeling, verbal features could be linguistic words for accepting like ok, hmm, yes, alright etc. non-verbal features could be like nodding head, eye blink.

List of categories and some of the features are described below:

Table 4. Verbal and non-verbal features for interaction.

Category	**Features**	
	Verbal	**Non-verbal**
Accept feeling	Yes, hmm, ok,	Nodding head, eye blink, hand-movement,
Praises or encourages	Good, fine, excellent, not bad, go on , continue, go ahead , tell us more	Hand clap, head nodding, smile, iris expansion, heptic
Accepts or uses ideas of students	Fine, ok, well that's interesting…, repeating student's statement.	Interaction through board, writing on board
Ask questions	Tell me, Why, how, when, where, what, is this, do you, board content	Pause, pointing to board,
Lectures	Explaining the topic	Hand movement, pointing to board, writing over board, maximized focus of class
Give directions	Stand-up, sit-down, keep quiet, raise your hand, calling by name	Pointing to student
Criticize or justifies authority	Listen, repeating own sentences,	
Pupil's response	Answering the question	Head nodding, writing, smile
Student talk-initiation	Asking question	Hand-raise, standing
Silence or confusion	Pause	

The above features may have some linguistic and Paralinguistic features, that can also be extracted with speech recognition and verbal features. Para-linguistic features represents phrases like hmm, exclamation, tone, pitch, loudness, intensity, speaker's state, mood, age, gender, personality, eye contact etc. Emotional state of a person can be determined by the linguistic and Paralinguistic features. For verbal feature extraction, natural language processing can be used to extract audio, speech and textual features like vocal tone, pitch, stress, energy related features. Some of the features are low-level descriptors (LLDs) including energy related LLDs, such as loudness, RMS energy, zero-crossing rates, spectral LLDs such as MFCC, spectral energy, spectral variance [32, 35, 36].

For extracting non-verbal features such as facial expression, face action units, hand-gestures, motion detection, body pose or gesture detection a set of computer vision based open source tools and techniques will be considered along with development of new tools wherever required.

For implementation, one or two classrooms will be equipped with multiple cameras and multiple audio devices to extract the video and audio data streams of the classroom interaction focusing on the teacher and the students. Some of the additional sensors will be installed to understand the classroom environment such as motion detection sensors, temperature, humidity and other gas sensors in the classroom. The recordings of about 100 sessions of classroom interactions taught by multiple teachers will be recorded for automatic analysis. One session is about 20 minutes. Classroom interaction will also be observed and analyzed by an expert faculty from the education faculty of the University for each session using standard interaction analysis models such as Flanders. For automation process, each 20 minutes recorded session is further spilt into multiple 3 sec short videos for interaction analysis to align with Flanders interaction model. Audio and Video analysis will be performed using available image and video processing open source tools. Development of new or enhancement of the available tools may be considered. For training and classification of audio and video, different machine learning models will be considered. To manage some of the challenges such as ambiguity in categorizing the interaction of a 3 sec slot or an ambiguous event of 10 sec silence, a fuzzy model based approach is also integrated into the system. As individual facial features and linguistics of teacher response may vary, the automated system may also be fine tuned to develop a personalized at a teacher level also. An automated interaction analysis feedback report is generated and submitted to the teacher to

understand the overall effectiveness of the interaction observed in the classroom. The human observer's and the automated systems outcome will be compared, evaluated and correlated accordingly.

The basic workflow of the methodology is described below:

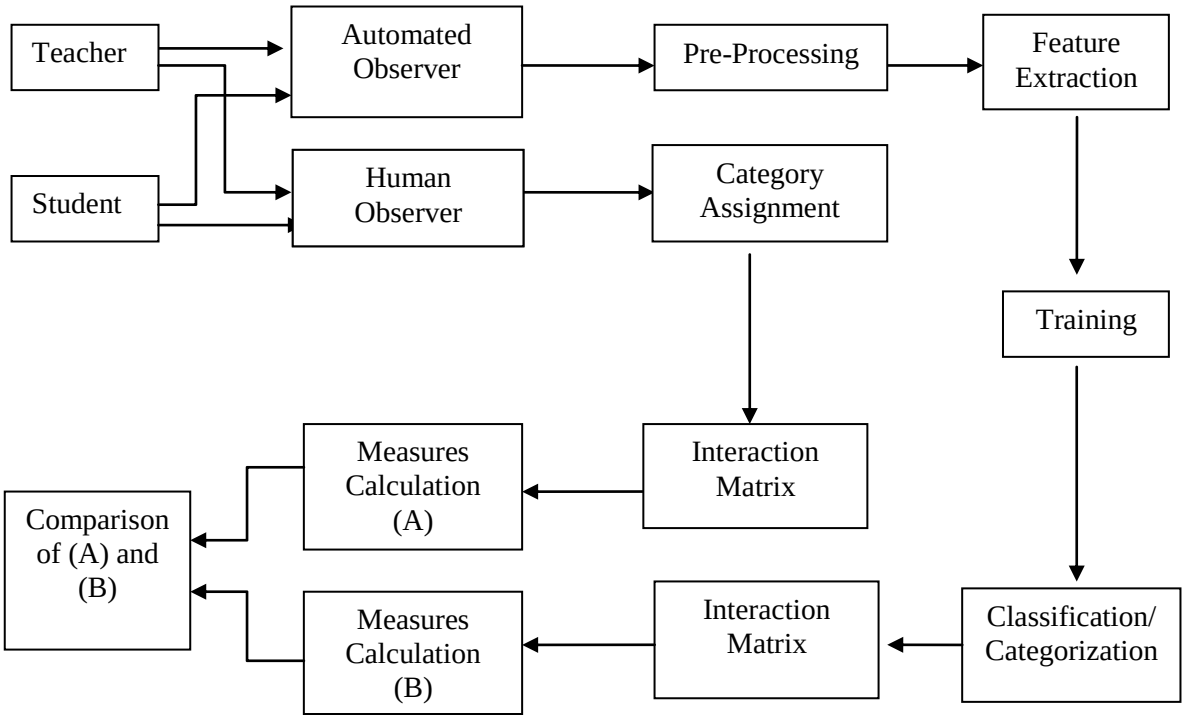

Figure 1. Basic workflow of methodology

4. Objectives

1. To study, explore and to conduct a qualitative survey on classroom interaction analysis based on verbal and non-verbal interaction.
2. To design and develop an automated classroom interaction analysis system based on verbal attributes inspired by Flanders interaction model using machine learning algorithms and to evaluate its performance by comparing with the human observer interaction analysis.
3. To design and develop an automated classroom interaction analysis system by integrating verbal and non-verbal interaction attributes inspired by Flanders' and Galloway's interaction models using machine learning algorithms and to evaluate its performance by comparing Flanders' verbal interaction model with Galloway's verbal and non-verbal model along with human observer interaction analysis.

4. To generate an interaction analysis feedback report to the teacher to represent the effectiveness of the interaction observed in the classroom.

References:

[1] Rohmah, Taufiqur. "Classroom Interaction in English Language Class for Students of Economics Education." *Arab World English Journal (AWEJ)* 8.2 (2017).

[2] Zhang, Jia-Hua, Liu-cong Zou, Jia-jia Miao, Ye-Xing Zhang, Gwo-Jen Hwang and Yue Zhu. "An individualized intervention approach to improving university students' learning performance and interactive behaviors in a blended learning environment." *Interactive Learning Environments* (2019): 1-15.

[3] Galloway, Charles. *"Silent Language in the Classroom. Fastback 86."* Phi Delta Kappa, Eighth & Union, Box 789, Bloomington, Indiana (1976).

[4] Flanders, Ned A. *"Analyzing teaching behavior."* Reading, Mass.: TheAddison-Wesley Publishing Company (1970).

[5] Amatari, Veronica Odiri. "The instructional process: a review of Flanders' interaction analysis in a classroom setting." *International Journal of Secondary Education* 3.5 (2015): 43-49.

[6] İnceçay, Görsev. "The role of teacher talk in young learners' language process." *Procedia-Social and Behavioral Sciences* 2.2 (2010): 277-281.

[7] Aisyah, Nurul. "An Analysis of Teachers' talk in an EFL Classroom." *Journal of English and Education* 4.2 (2016): 63-79.

[8] Obaid, Mohammad, Ruth Aylett, Wolmet Barendregt, Christina Basedow, Lee J. Corrigan, Lynne Hall, Aidan Jones, Arvid Kappas, Dennis Kuster, Ana Paiva, Fotis Papadopoulos, Sofia Serholt, Ginevra Castellano1. "Endowing a robotic tutor with empathic qualities: Design and pilot evaluation." *International Journal of Humanoid Robotics* 15.06 (2018): 1850025.

[9] Kim, Yelin, Tolga Soyata, and Reza Feyzi Behnagh. "Towards emotionally aware AI smart classroom: Current issues and directions for engineering and education." *IEEE Access* 6 (2018): 5308-5331.

[10] Spikol, Daniel, Luis P. Prieto, and M.J. Rodriguez-Triana. "Current and future multimodal learning analytics data challenges." *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. ACM, 2017.

[11] Blikstein, Paulo. "Multimodal learning analytics." *Proceedings of the third international conference on learning analytics and knowledge*. ACM, 2013.

[12] Prieto, Luis Pablo, Kshitij Sharma, Łukasz Kidzinski, María Jesús Rodríguez-Triana, and Pierre Dillenbourg. "Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data." *Journal of computer assisted learning* 34.2 (2018): 193-203.

[13] Nerina Reitano. *"Digital Emotions : the potential and issues of Affective Computing systems ". University of Dublin, 2018.*

[14] Donnelly, Patrick J., Nathan Blanchard, Borhan Samei, Andrew M. Olney, Xiaoyi Sun, Brooke Ward, Sean Kelly, Martin Nystrand, and Sidney K. D'Mello. "Automatic teacher modeling from live classroom audio." *Proceedings of the 2016 conference on user modeling adaptation and personalization*. ACM, 2016.

[15] Donnelly, Patrick J., Nathaniel Blanchard , Andrew M. Olney , Sean Kelly , Martin Nystrand , and Sidney K. D'Mello1. "Words matter: automatic detection of teacher questions in live classroom discourse using linguistics, acoustics, and context." *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. ACM, 2017.

[16] Boz, Hakan, and Utku Kose. "Emotion Extraction from Facial Expressions by Using Artificial Intelligence Techniques." *BRAIN. Broad Research in Artificial Intelligence and Neuroscience* 9.1 (2018): 5-16.

[17] Singh, Y. K., T. K. Sharma, and Brijesh Upadhya. *Educational Technology: Teaching-learning.* APH, 2008.

[18] Domínguez, Federico, Vanessa Echeverría, Katherine Chiluiza and Xavier Ochoa. "Multimodal selfies: Designing a multimodal recording device for students in traditional classrooms." *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 2015.

[19] Worsley, Marcelo, and R. Martinez-Maldonado. "Multimodal Learning Analytics' Past, Present, and Potential Futures." *CrossMMLA@ LAK*. 2018.

[20] Hildmann, Hanno. "Designing Behavioural Artificial Intelligence to Record, Assess and Evaluate Human Behaviour." *Multimodal Technologies and Interaction* 2.4 (2018): 63.

[21] Geiger, Jürgen T., Björn Schuller, and Gerhard Rigoll. "Large-scale audio feature extraction and SVM for acoustic scene classification." *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2013.

[22] Sapp, Ben, and Ben Taskar. "Modec: Multimodal decomposable models for human pose estimation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013.

[23] Li, Ya, Jianhua Tao, Björn Schuller, Shiguang Shan, Dongmei Jiang and Jia Jia. "Mec 2017: Multimodal emotion recognition challenge." *2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*. IEEE, 2018.

[24] El Ayadi, Moataz, Mohamed S. Kamel, and Fakhri Karray. "Survey on speech emotion recognition: Features, classification schemes, and databases." *Pattern Recognition* 44.3 (2011): 572-587.

[25] Andrade, Alejandro, Ginette Delandshere, and Joshua A. Danish. "Using multimodal learning analytics to model student behavior: A systematic analysis of epistemological framing." *Journal of Learning Analytics* 3.2 (2016): 282-306.

[26] Kucherenko, Taras, Dai Hasegawa, Gustav Eje Henter, Naoshi Kaneko, and Hedvig Kjellström. "Analyzing Input and Output Representations for Speech-Driven Gesture Generation." *arXiv preprint arXiv:1903.03369* (2019).

[27] Takeuchi, Kenta, Souichirou Kubota, Keisuke Suzuki, Dai Hasegawa and Hiroshi Sakuta. "Creating a gesture-speech dataset for speech-based automatic gesture generation." *International Conference on Human-Computer Interaction*. Springer, Cham, 2017.

[28] Dillenbourg, Pierre, Sanna Järvelä, and Frank Fischer. "The evolution of research on computer-supported collaborative learning." *Technology-enhanced learning*. Springer, Dordrecht, 2009. 3-19.

[29] Littlewort, Gwen, Marian Bartlett, Tingfan Wu and Javier Movellan. "The computer expression recognition toolbox (CERT)." Face and gesture 2011. IEEE, 2011.

[30] Giannakopoulos, Theodoros. "pyaudioanalysis: An open-source python library for audio signal analysis." *PloS one* 10.12 (2015): e0144610.

[31] Yao, Qingmei. *"Multi-sensory Emotion recognition with speech and facial expression"* (2014). University of Southern Mississippi. Dissertations. https://aquila.usm.edu/dissertations/710

[32] Schuller, Björn, Stefan Steidl, Anton Batliner, Felix Burkhardt, Laurence Devillers, Christan Mullar and Shrikanth Narayanan. "Paralinguistics in speech and language—State-of-the-art and the challenge." *Computer Speech & Language* 27.1 (2013): 4-39.

[33] Schuller, Björn, Anton Batliner, Stefan Steidl and Dino Seppi. "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge." *Speech Communication* 53.9-10 (2011): 1062-1087.

[34] Eyben, Florian, Martin Wöllmer, and Björn Schuller. "Opensmile: the munich versatile and fast open-source audio feature extractor." *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010.

[35] Schuller, Björn, Stefan Stedl, Anton Batliner, Alessandro Vinciarelli, Klaus Scherer, Fabien Ringeval, Mohamed Chetouani, Felix Weninger, Florian Eyben, Erik Marchi, Marcello Mortillaro, Hugues Salamin, Anna Polychroniou, Fabio Valente and Samuel Kim. "The Interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism." *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France*. 2013.

[36] Schuller, Björn, Anton Batliner, Stefan Steidl Elmar Noth, Alessandro Vinciarelli, Felix Burkhardt Rob van Son, Felix Weninger, Florian Eyben, Tobias Bocklet, Gelareh Mohammadi, Benjamin Weiss. "The Interspeech 2012 speaker trait challenge." *Thirteenth Annual Conference of the International Speech Communication Association*. 2012.

[37] Ängeslevä, Jussi, Carson Reynolds, and Sile O'Modhrain. "EmoteMail." *ACM SIGGRAPH 2004 Posters*. ACM, 2004.

[38] Galloway, Charles. "Nonverbal communication." *Theory into practice* 7.5 (1968): 172-175.

[39] Galloway, Charles M. "Teaching is communicating: Nonverbal Language in the Classroom." Washington, D. C. ,Association for Student Teaching, NEA, Bulletin No. 29. (1970).

[40] Galloway, Charles. "Nonverbal Communication in The Classroom." Distributed by ERIC Clearinghouse [Washington, D.C.] (1967). https://files.eric.ed.gov/fulltext/ED011891.pdf

[41] Ellis, Roger. "Simulated social skill training for interpersonal professions." *The analysis of social skill*. Springer, Boston, MA, 1980. 79-101.

[42] Davis, James. *Teaching strategies for the college classroom*. Routledge, 2019.

[43] George, A. D., and Phil Brown. *Microteaching: A Programme of Teaching Skills*. Methuen, 1975.

[44] India, Ministry of Human Resource Development. "Draft National Education Policy 2019." https://mhrd.gov.in/sites/upload_files/mhrd/files/Draft_NEP_2019_EN_Revised.pdf